AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2016WR019024

# Precipitation estimation using L-band and C-band soil moisture retrievals

**Randal D. Koster[1], Luca Brocca[2], Wade T. Crow[3], Mariko S. Burgin[4], and Gabrielle J. M. De Lannoy[5]**

[1]Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, Maryland, USA, [2]Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy, [3]USDA Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA, [4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA, [5]Department of Earth and Environmental Sciences, KU Leuven, Heverlee, Belgium

**Key Points:**
- The time sequencing of rain is extracted from 4 soil moisture retrieval data sets
- This rain estimation is of unprecedented accuracy for the L-band retrievals
- Known features of the instruments and algorithms explain their relative performance

**Correspondence to:**
R. Koster,
randal.d.koster@nasa.gov

**Citation:**
Koster, R. D., L. Brocca, W. T. Crow, M. S. Burgin, and G. J. M. De Lannoy (2016), Precipitation estimation using L-band and C-band soil moisture retrievals, *Water Resour. Res.*, *52*, doi:10.1002/2016WR019024.

Received 4 APR 2016
Accepted 1 SEP 2016
Accepted article online 6 SEP 2016

**Abstract** An established methodology for estimating precipitation amounts from satellite-based soil moisture retrievals is applied to L-band products from the Soil Moisture Active Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS) satellite missions and to a C-band product from the Advanced Scatterometer (ASCAT) mission. The precipitation estimates so obtained are evaluated against in situ (gauge-based) precipitation observations from across the globe. The precipitation estimation skill achieved using the L-band SMAP and SMOS data sets is higher than that obtained with the C-band product, as might be expected given that L-band is sensitive to a thicker layer of soil and thereby provides more information on the response of soil moisture to precipitation. The square of the correlation coefficient between the SMAP-based precipitation estimates and the observations (for aggregations to ~100 km and 5 days) is on average about 0.6 in areas of high rain gauge density. Satellite missions specifically designed to monitor soil moisture thus do provide significant information on precipitation variability, information that could contribute to efforts in global precipitation estimation.

## 1. Introduction

The potential societal benefits of an accurate estimation of precipitation—its magnitudes and its variations in time—are immense. Precipitation data are crucial for crop modeling and forecasting, water resources planning, soil moisture initialization for weather forecasts and seasonal forecasts, flood and landslide analysis, and a host of other valuable applications. Precipitation is indeed the key driver of surface hydrological processes and is an essential link between the land and atmospheric components of the climate system.

The importance of measuring precipitation accurately has not been lost on the scientific community. A number of projects over the years have produced global-scale precipitation data for scientific and technical applications. Key data sets are available, for example, from the National Center for Environmental Prediction [*Xie et al.*, 2007, also ftp://ftp.cpc.ncep.noaa.gov/precip/cmap/] and the Global Precipitation Climatology Project [*Adler et al.*, 2003], the latter being sponsored by the Global Climate Research Programme. To produce the global-scale gridded precipitation rates, such projects utilize a number of data sources, including rain gauges, satellite-based precipitation retrievals, and model analysis products. Satellite-based estimates of precipitation are indeed becoming more and more relevant, with valuable data provided by the Tropical Rain Measurement Mission (TRMM) [*Huffman et al.*, 2007] and the follow-on Global Precipitation Mission (http://www.nasa.gov/mission_pages/GPM/main/index.html).

Advances in technologies notwithstanding, all current precipitation estimation techniques have limitations. Rain gauges are generally considered to be the most accurate source of precipitation data [*Huffman et al.*, 1997], but they represent local measurements and, given issues of spatial representativeness, are not always easily translated to area-averaged precipitation rates. Rain gauges are also absent in many parts of the world. Satellite-based precipitation rates, while clearly valuable, are limited by the "snapshot" character of the individual measurements and by various difficulties in interpreting satellite signals over land [*Hou et al.*, 2014; *Kummerow et al.*, 2015]. Analysis products from atmospheric models, for their part, are subject to the biases inherent in the underlying models used. Given such issues, alternative methods of estimating

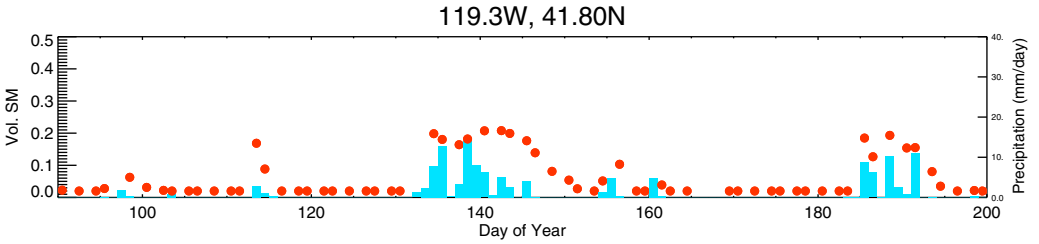


**Figure 1.** SMAP soil moisture retrievals (red dots) at a dry western US location plotted alongside contemporaneous rain gauge-based precipitation data at the same location (blue histogram bars). Soil moisture is in volumetric units, and precipitation is in mm/d. The site, in northwestern Nevada, is characterized by scrubby vegetation.

precipitation could prove valuable. Indeed, using data from a proven alternative method in concert with satellite-based precipitation retrievals, rain gauge measurements, and analysis products, properly taking into account the relative strengths and weaknesses of each method, could yield a superior global precipitation data set that could benefit many user applications.

One particularly promising and currently under-utilized data source relevant to precipitation estimation is soil moisture as measured from space. The potential for extracting precipitation information from space-based soil moisture retrievals is illustrated in Figure 1, which shows time series of Level 2 passive soil moisture retrievals from the Soil Moisture Active-Passive mission (SMAP; see section 2.1.1) plotted alongside spatially collocated gauge-based precipitation data at a western U.S. site. SMAP surface soil moisture values (top ~5 cm of soil) are represented as red dots, and the precipitation rates, from the Climate Prediction Center Unified rain gauge data set (see section 2.1.2), are shown as blue histogram bars. Soil moisture is seen to increase at the onset of precipitation (e.g., on days 117, 133, 155, and 184), with larger increases for larger precipitation rates (compare the increases on days 133 and 155). Furthermore, following the cessation of rain, soil moisture gradually reduces to a value near zero. The overall consistency between the independent soil moisture and precipitation data is high; the retrievals here do contain useful information on the time sequencing and relative magnitudes of precipitation events.

Recognizing this connection, several studies have in fact presented approaches for utilizing satellite-based soil moisture data to improve existing precipitation data sets [e.g., *Crow et al.*, 2009, 2011; *Pellarin et al.*, 2008, 2013; *Wanders et al.*, 2015; *Zhan et al.*, 2015]. In the present study, we consider a distinctly different class of algorithm: one that uses soil moisture retrievals in isolation to compute an independent time series of precipitation for direct comparison with existing data from more traditional sources (e.g., rain gauges or satellites). This use of soil moisture as a virtual rain gauge was pioneered by *Brocca et al.* [2013], who developed a specific algorithm, called SM2RAIN, for generating time series of precipitation rates based on the changes seen in consecutive soil moisture retrievals (see section 2.1.3). *Brocca et al.* [2014] applied this algorithm to Advanced Scatterometer (ASCAT) and Advanced Microwave Scanning Radiometer—EOS (AMSR-E) soil moisture retrievals and to an early version of Soil Moisture and Ocean Salinity (SMOS) retrievals and found that the resulting precipitation time series were promisingly realistic.

The launch in 2015 of the SMAP soil moisture satellite and the considerable updates in the processing of the SMOS products (since the *Brocca et al.* [2014] study) provides a valuable opportunity to evaluate this precipitation estimation approach with presumably more accurate soil moisture retrievals. SMAP and SMOS are both L-band instruments and thereby extract soil moisture information from deeper in the soil than C-band instruments such as AMSR-E or ASCAT (~5 cm versus ~2 cm), allowing for a more complete characterization of how soil moisture responds to precipitation. The SMAP mission, in addition, has numerous protocols in the place to reduce noise from radio frequency interference (RFI) [*Entekhabi et al.*, 2010]. The present paper aims to quantify the level of precipitation estimation accuracy achievable using these new L-band instruments relative to that achievable with a representative C-band instrument.

Section 2 describes the data sets used and outlines the SM2RAIN precipitation estimation algorithm, including special modifications adopted for this study. Section 3 presents the accuracies achieved with the L-band and C-band data, and section 4 provides a summary and further discussion.

## 2. Data and Methods

### 2.1. Satellite Retrievals

#### 2.1.1. SMAP

The SMAP satellite, launched in early 2015, carries an L-band radar and radiometer that provide global radar backscatter and brightness temperature observations every 2–3 days. The radar ceased operation on 7 July 2015, but the radiometer continues to operate well. Among other products, SMAP retrieves the soil moisture content of the upper ~5 cm of soil. SMAP was designed with a sun-synchronous orbit with 6 AM/PM local equatorial overpass time and has a nominal incidence angle of 40°.

The specific SMAP data used in this study are the Level 2 retrievals (L2_SM_P) from the passive radiometer [*Entekhabi et al*., 2010; *Chan and Dunbar*, 2015]. The passive-based soil moisture data are provided on a 36 km Earth-fixed grid using the global cylindrical Equal-Area Scalable Earth Grid Projection Version 2 (EASEv2) [*Brodzik et al*., 2012]. We use in particular the "beta release" version of these data [*O'Neill et al*., 2014], the most advanced version available to us at the time of this writing. These data are limited to the descending swaths of the SMAP data, corresponding to local retrieval times of 6 AM. The retrievals are obtained using the Single Channel Algorithm [*Jackson et al*., 2015] and are currently based on V-polarized brightness temperature observations only. We excluded coastal pixels from our analysis, and we considered only those retrievals that have been flagged as "attempted" and "successful." To allow greater global coverage, however, we ignored the flag associated with "recommended quality." We also ignored flags indicating the potential presence of snow or frozen soil; given the time period considered (May for calibration, and mid-June through mid-October for validation, as discussed below), this should have minimal impact on our results over most of the globe.

#### 2.1.2. SMOS-A and SMOS-D

SMOS [*Kerr et al*., 2010], launched in early November 2009, carries an L-band radiometer and primarily maps soil moisture and ocean salinity. It observes the Earth in a sun-synchronous orbit at 6 AM/PM local overpass time at incidence angles ranging from 0° to 65°, and, like SMAP, it has a temporal revisit of 2–3 days and a nominal spatial resolution of about 40 km. For this study, we use Level 2 retrieval data from the SMOS SMUDP2 product version v620. The SMOS retrieval algorithm simultaneously retrieves soil moisture and other variables, such as the vegetation opacity, by fitting multiangular brightness temperatures at both horizontal and vertical polarization with L-band Microwave Emission of the Biosphere (L-MEB) [*Wigneron et al*., 2007] model simulations. Data were retained only if: (a) all retrieved variables fall within a realistic range (0–0.8 $m^3/m^3$ for soil moisture), (b) the retrieval uncertainty is less than a certain threshold (0.1 $m^3/m^3$ for soil moisture), (c) the RFI-probability for both H-polarization and V-polarization is less than 0.3, and (d) flags are not raised for high topographic complexity, high urban fraction, high open water fraction, sea ice, coastal areas, and high total electron content. The SMOS data were regridded from a 12.5 km posting resolution to the 36 km EASEv2 grid; during this aggregation step, the data were screened for excessive sub-36 km heterogeneity (spatial standard deviation $> 0.5$ $m^3/m^3$) that may be indicative of RFI or the presence of open water bodies.

We use two distinct SMOS data sets in this study: SMOS-A, consisting of data collected on ascending passes of the satellite (corresponding to 6 AM local time), and SMOS-D, consisting of data collected on descending passes (corresponding to 6 PM local time). The data are separated in this way because the timing of the overpass has a potentially significant impact on retrieval accuracy [see, e.g., *Lei et al*., 2015]. By using both SMOS data sets, we should be able to see if the expected increase in accuracy for SMOS-A translates to a corresponding increase in the accuracy of precipitation estimation.

#### 2.1.3. ASCAT

ASCAT, a real aperture radar operating at C-band, was launched on board the European Meteorological Operational (MetOp)-B spacecraft in 2012. It observes the Earth in a sun-synchronous orbit at 9:30 AM/PM local overpass time, and it has a temporal revisit of 3 days. For this study, we took advantage of the availability of an ASCAT data set already processed by the SMAP mission for comparison with SMAP morning retrievals. To construct this data set, the 9:30 AM (descending) ASCAT L2 soil moisture index posted at 12.5 km resolution was regridded to EASEv2 at 36 km by averaging the data using inverse distance weighting for each day. ASCAT retrievals were masked out if the probability of snow, frozen ground, wetland, or significant topography exceeds 50% or if the soil moisture estimation uncertainty due to other sources exceeds 50%. The soil moisture index on EASEv2 at 36 km was converted to volumetric soil moisture by

multiplication with soil porosity, which was also delivered (at 9 km) as ancillary data [*De Lannoy et al.*, 2014; *Mahanama et al.*, 2015] to the SMAP Level 4 soil moisture product [*Entekhabi et al.*, 2014].

### 2.2. Precipitation Data

The precipitation data used to evaluate the satellite-based precipitation estimates are from the Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Global Daily Precipitation (hereafter CPCU; see ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/). For this study, the daily data set was converted from its original 0.5° × 0.5° grid to the SMAP EASEv2 grid at 36 km using areal weighting.

As its name implies, the CPCU data are based on rain gauges only; no satellite-based rainfall information was used in the construction of the data set. In focusing on the gauge-based data, we implicitly assume that it is the most accurate data available. Indeed, gauge-based data are generally used to validate satellite-based precipitation retrievals [*Huffman et al.*, 1997]. The usefulness of the data set for validation is nevertheless limited in regions of low rain gauge density. Figure 2 shows the rain gauge density associated with the CPCU data used. High densities are seen in much of North America and Europe and in various parts of the other continents. On the other hand, low densities appear, for example, in high northern latitudes, in the Amazon, and in most of Africa. In such low-density regions, we cannot pretend to know (from the CPCU data set or, arguably, from gauge-based precipitation data sets in general) what the true daily precipitation rates are. We will refer to the density map in Figure 2 as we proceed with our analyses. Note that the map uses density units of #gauges/0.5°×0.5°-cell; these densities, with no change in units, are regridded using

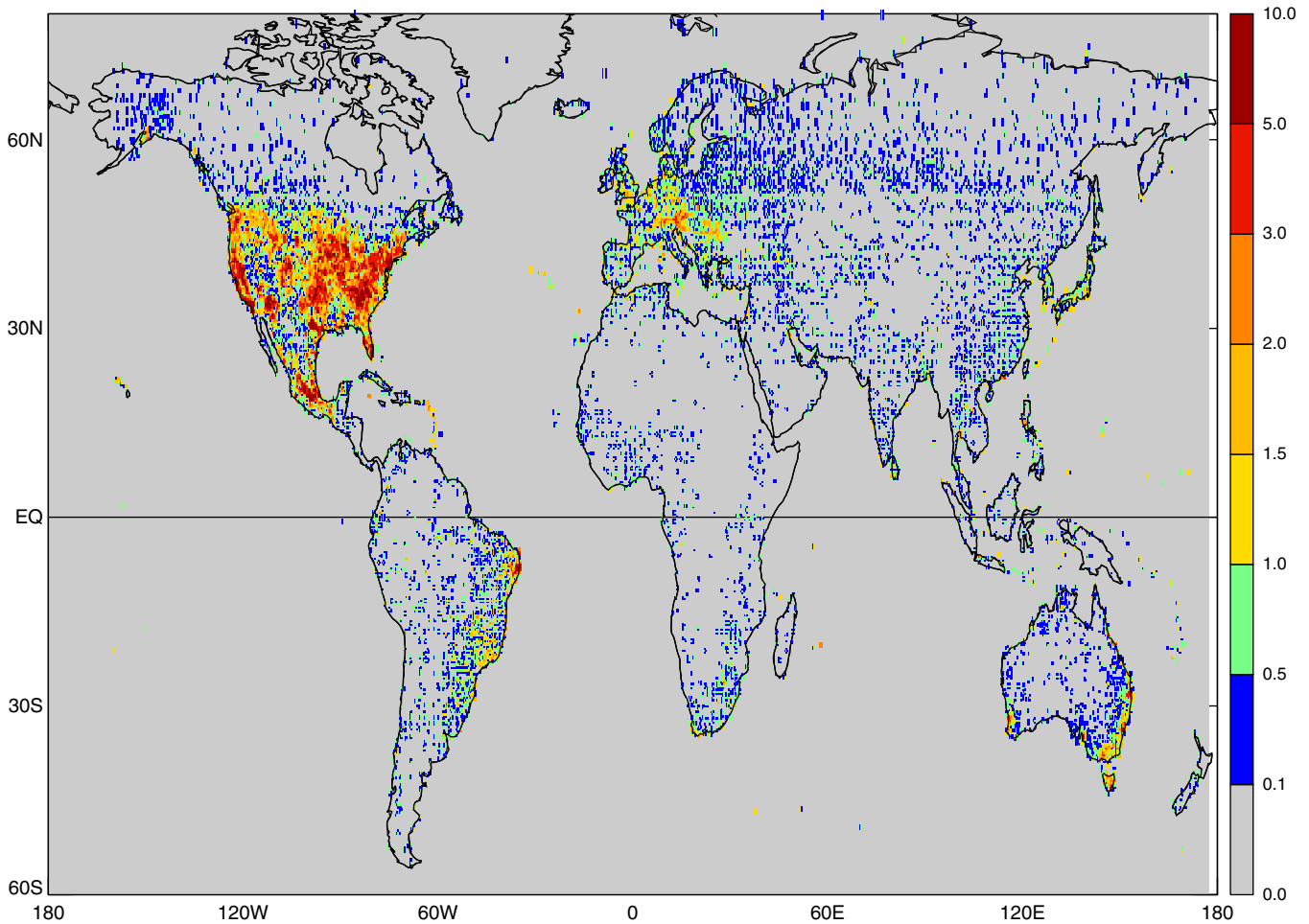


**Figure 2.** Density of rain gauges underlying the CPCU precipitation data set used to evaluate the soil moisture retrieval-based precipitation estimates. Data were provided by CPCU in units of #gauges/0.5°×0.5° grid cell; the data were translated to the SMAP EASEv2 grid at 36 km while retaining the original units.

areal weighting to the finer EASEv2 grid at 36 km for use in evaluating our precipitation estimation accuracy.

Another important issue regarding the CPCU precipitation data involves the reporting time for the daily values, which differs by region—some regions may report values for 6 AM-6 AM local time to the CPCU data collectors, others may report calendar-day values, and so on. To reduce the impact of the potential inconsistency between the gauge precipitation measurements and the retrieval-based estimates, we will focus our validation on 5 day precipitation totals; for each day in the validation period, we compare the estimated total precipitation from 2 days prior to 2 days after the reported date to the corresponding total from CPCU. Through such a procedure, of course, some inconsistency may still remain on day −2 and day +2. Note that this remaining inconsistency can only reduce the computed precipitation estimation skill levels, so that true skill levels may in fact be higher than those established here.

Finally, we do not attempt here to separate the observed precipitation rates into rainfall and snowfall rates. Again, given the time period considered in this analysis (northern hemisphere warm season), this should have limited impact on our results over most of the globe.

### 2.3. The SM2RAIN Precipitation Estimation Algorithm

In its basic form, the SM2RAIN algorithm [*Brocca et al*., 2013] estimates the precipitation, $P_{est}$, for each day between retrieval times t−1 and t using an equation equivalent to:

$$P_{est} = \alpha \ \mathrm{Max}\{0., \ [(W_t - W_{t-1})/\Delta t + 0.5 \ \gamma \ (W_t^{\,b} + W_{t-1}^{\,b})]\} \tag{1}$$

where $W_t$ and $W_{t-1}$ are the consecutive soil moisture retrievals (in volumetric units: $m^3/m^3$), $\Delta t$ is the time interval (in days) between them, and $\alpha$, $\gamma$, and b are the fitted constants (see below), with $\gamma$ having units that convert the second term within the brackets of equation (1) to units of $m^3/m^3/d$, and $\alpha$ having units that convert the right-hand side of equation (1) to mm/d. The term $(W_t - W_{t-1})/\Delta t$ is positive if soil moisture increases between t−1 and t; this increase is indicative of a precipitation event and thus adds to the value of $P_{est}$. The term $0.5\gamma \ (W_t^b + W_{t-1}^b)$ is included to represent loss (e.g., drainage), which can reduce surface soil moisture even during precipitation events. Because this drainage is larger for wetter soils, precipitation has to "fight harder" to increase soil moisture when the soil is wetter; this second term captures this effect. The presence of this term allows (1) to estimate nonzero rainfall even when the soil moisture decreases slightly over the time interval. (Note that the original equation in *Brocca et al*. [2013] only included the $\gamma \ W_t^b$ term; here, a second term is included to tie the assumed drainage to both the initial and final soil moisture states to approximate an average drainage.)

Of course, any such algorithm has an important limitation: its ability to capture high precipitation rates is necessarily limited by the fact that soil moisture cannot exceed porosity, so that any precipitation water that forms overland flow will necessarily be missed. Also, the imprint of a given precipitation volume on a soil moisture retrieval will presumably depend on how long before the retrieval the precipitation event occurred, and satellite retrievals in any case contain error that will necessarily be propagated to the precipitation estimates. Even so, *Brocca et al*. [2014] demonstrate a successful application of the algorithm to ASCAT data, and, as will be shown in the following section, the algorithm performs even better with SMAP and SMOS data.

### 2.4. Application of SM2RAIN to SMAP, SMOS, and ASCAT Data

#### 2.4.1. Skill Metric

Our metric for evaluating the algorithm's ability to estimate precipitation is the square of the correlation coefficient ($r^2$) between our precipitation estimates from equation (1) and corresponding observed (gauge-based) precipitation rates. Thus, in this paper, we are evaluating the estimation of the time sequencing of precipitation and the associated capture of the relative magnitudes of different storms rather than the absolute magnitudes of the rates, as would be addressed with a root-mean-square-error metric. By using an $r^2$ metric, we are in fact evaluating a quantity that is directly proportional to the actual precipitation rate, which has the distinct advantage of reducing from 3 to 2 the number of parameters needing calibration in equation (1)—there is no need to calibrate the scale factor $\alpha$. When it comes time to producing actual precipitation estimates, our estimates would need to be scaled accordingly, presumably in a very simple way using ratios of long-term observed precipitation totals to long-term estimate totals, either in the region of

interest or, for a region without adequate precipitation measurement, in a region of similar soil texture. Alternatively, the information contained in the (unscaled) time sequences could be used directly in conjunction with other precipitation time series (e.g., from rain gauges, satellite missions focused on rainfall) to produce improved hybrid data sets—a distinct possibility if the soil moisture-based information is determined to be significant through the $r^2$ metric.

The satellite soil moisture retrievals are not available on a daily basis; they are often separated by 2 or 3 days. The effective temporal resolution of the associated SM2RAIN precipitation estimates is necessarily tied to these retrieval times. In our analyses, if two consecutive retrieval intervals are separated by N days, the resulting SM2RAIN estimate of precipitation rate from (1) is assigned to each of those N days. As noted in section 2.2, we further coarsen the resulting daily time series of estimated precipitation rates to a sequence of 5 day averages, which we compare to corresponding 5 day averages of rain gauge data from the CPCU data set.

#### 2.4.2. Special Modifications of the Basic Algorithm

In practice, different sets of values for the parameters in (1) can be determined for different regions of the world. *Brocca et al*. [2014] indeed use different climatic precipitation classes to define different parameter sets. For this study, however, we emphasize simplicity and robustness; we determine a single set of parameters that can be used everywhere across the globe. Going to region-specific or hydrological regime-specific parameter sets would theoretically only increase our computed estimation accuracies.

Using a single set of parameters makes it necessary, when processing the satellite retrievals, to standardize soil moisture contents by: (i) determining, at each grid element, the minimum soil moisture obtained over the period of record, and then (ii) subtracting this value from each retrieval at that grid element. In conceptual terms, such a calculation has both an advantage and a disadvantage. The advantage is that it addresses the fact that different locations on the globe may (at least for certain retrieval data sets) have different soil moisture minima, as a function, for example, of soil texture. The subtraction in effect provides all locations with a single common baseline—any soil moisture above the baseline, anywhere across the globe, has the potential to decrease during an interstorm period. The disadvantage is that many locations may never experience their true minimum value during the period of record, so that the baseline utilized for them is inaccurate. We proceed with full knowledge of this disadvantage, knowing that the inaccuracy would eventually be reduced as more satellite data are collected, and furthermore realizing that the inaccuracy, as it currently exists, can only degrade the performance of our present calculations. If the precipitation algorithm is shown now to perform well despite the inaccuracy in the estimated baseline soil moisture, then the inaccuracy can be assumed unimportant.

#### 2.4.3. Algorithm Calibration and Validation

For each of the satellite-based soil moisture retrieval data sets (SMAP, SMOS-D, SMOS-A, and ASCAT), we use the period 5–31 May 2015 (a period defined by mutual data availability) to calibrate the parameters $\gamma$ and $b$ in (1). Because we are using an $r^2$ metric, an arbitrary value for the parameter $\alpha$ is assigned. Our calibration procedure involves computing May precipitation time series using (1) for each of a great many potential pairings in the [$\gamma$, b] parameter space and determining the pairing for which the global average of the $r^2$ skill metric for May (in regions with a rain gauge density of at least 1 gauge per 0.5° grid cell) is the largest. The calibrated parameter values are then used to estimate precipitation over the period 20 June to 15 October. (Part of June is skipped in accordance with the European Space Agency's recommendation to avoid this particular data period for SMOS; see https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/smos/news/-/article/smos-level-1-and-2-data-products-short-period-of-degraded-data.)

## 3. Results

The red dots in Figure 3a show the time series of SMAP soil moisture retrievals at a representative location in the central US (a farmland region in southwestern Kansas). The dark blue histogram bars in Figure 3a show the precipitation time series estimated from these retrievals using (1). Careful study of these data shows that nonzero precipitation is indeed sometimes estimated with (1) even during periods of decreasing soil moisture, especially when the initial soil moisture is high.

Figure 3b shows, with dark blue hollow histogram bars, the corresponding time series of the SM2RAIN estimates averaged over 5 day periods. Plotted as light blue solid histogram bars are the observed 5 day

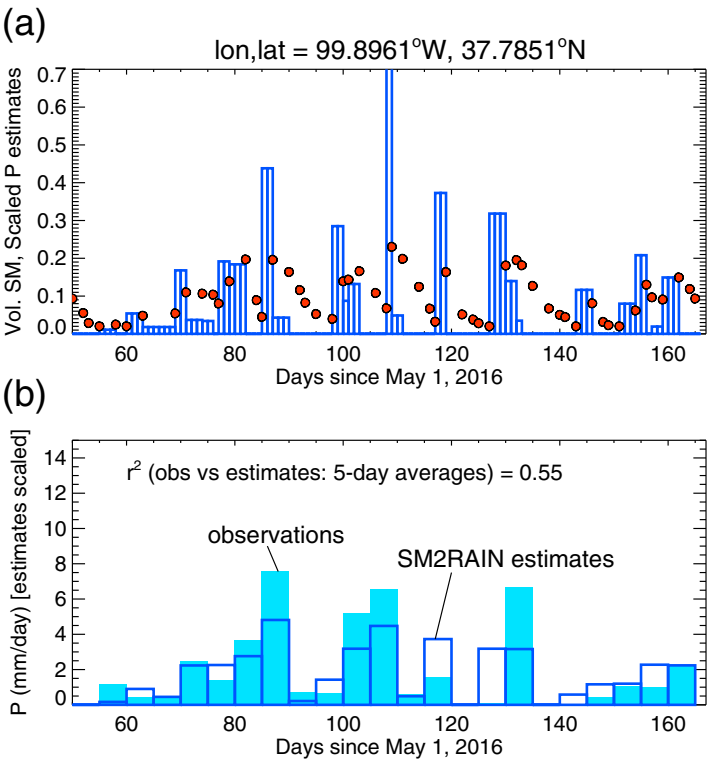


**Figure 3.** (a) Time series of SMAP soil moisture retrievals (red dots) at a representative location in the central US. Plotted with dark blue and hollow histogram bars are the daily precipitation rates estimated from these retrievals with the SM2RAIN algorithm, using (1). Note that if consecutive retrievals are separated by, say, 3 days, the resulting single precipitation estimate is assigned to each of the intervening 3 days. (b) Five day averages of the SM2RAIN precipitation estimates (dark blue and hollow bars) and corresponding 5 day totals from rain gauges at the same location (light blue solid histogram bars). For display purposes, the SM2RAIN estimates are arbitrarily scaled by a constant factor in each plot.

precipitation totals from the CPCU rain gauge data set. The time series show some similarity; the $r^2$ between the estimated and observed time series in Figure 3b is 0.55, indicating some skill in the estimation of precipitation from the soil moisture retrievals alone—over 50% of the observed precipitation variance is explained by our precipitation estimates.

This basic calculation underlies Figures 4a–4d, which show the global distributions of $r^2$ for 5 day precipitation rates estimated from the SMAP, SMOS-A, SMOS-D, and ASCAT data sets, respectively. Note that at some locations (shown in white), $r^2$ values could not be determined due to limitations in the precipitation data or in the soil moisture retrieval data (e.g., high levels of RFI). High $r^2$ values (exceeding 0.6) are seen, for example, in much of the continental United States and Europe and in parts of western Asia, Australia, southern Africa, and South America, particularly for SMAP. Lower $r^2$ values are seen elsewhere, but these do not necessarily imply a deficiency in the technique—rather, they are at least partly indicative of deficiencies in the precipitation observations (i.e., the validation data) themselves. This can be seen by comparing the fields in Figure 4 with the map of rain gauge density in Figure 2. The $r^2$ fields are strongly determined by rain gauge density, with high $r^2$ values generally found in regions of high density and low values in regions of low or zero density. We can reasonably argue that the true precipitation is simply not well known in areas with low gauge density and that, if the true precipitation were in fact known better in these regions, the skill found for the satellite-based estimates there would be much larger.

We can increase skill levels further by addressing spatial representativeness error. As described above, averaging the daily precipitation estimates to 5 day totals allows us to address some of the representativeness errors associated with inconsistencies in the timing of satellite overpasses and precipitation rain gauge measurements. Some representativeness errors, however, also exist in the spatial domain—rain gauges provide point measurements that may be inconsistent with the areal averages computed with the estimation algorithm, particularly in areas of lower gauge density. Furthermore, while the nominal (3 dB) resolution of, for example, SMAP and SMOS is ~40 km, the integrated signal in fact comes from a circular area with a diameter of ~80 km, with less weight in the outer area. To address (at least to some extent) these issues, we now compute correlations after aggregating both the retrieval-based precipitation estimates and the gridded rain gauge measurements to a coarser (~100 km, or about 1°) spatial scale: over 3 × 3 blocks of EASEv2 grid cells.

Figure 5 shows the results for all four retrieval data sets. The increase in the $r^2$ values is striking. As expected, values are still low in areas of low rain gauge density (as presumably they must be), but $r^2$ values are high across much of North America, Europe, and western Asia and are also high in many parts of the other

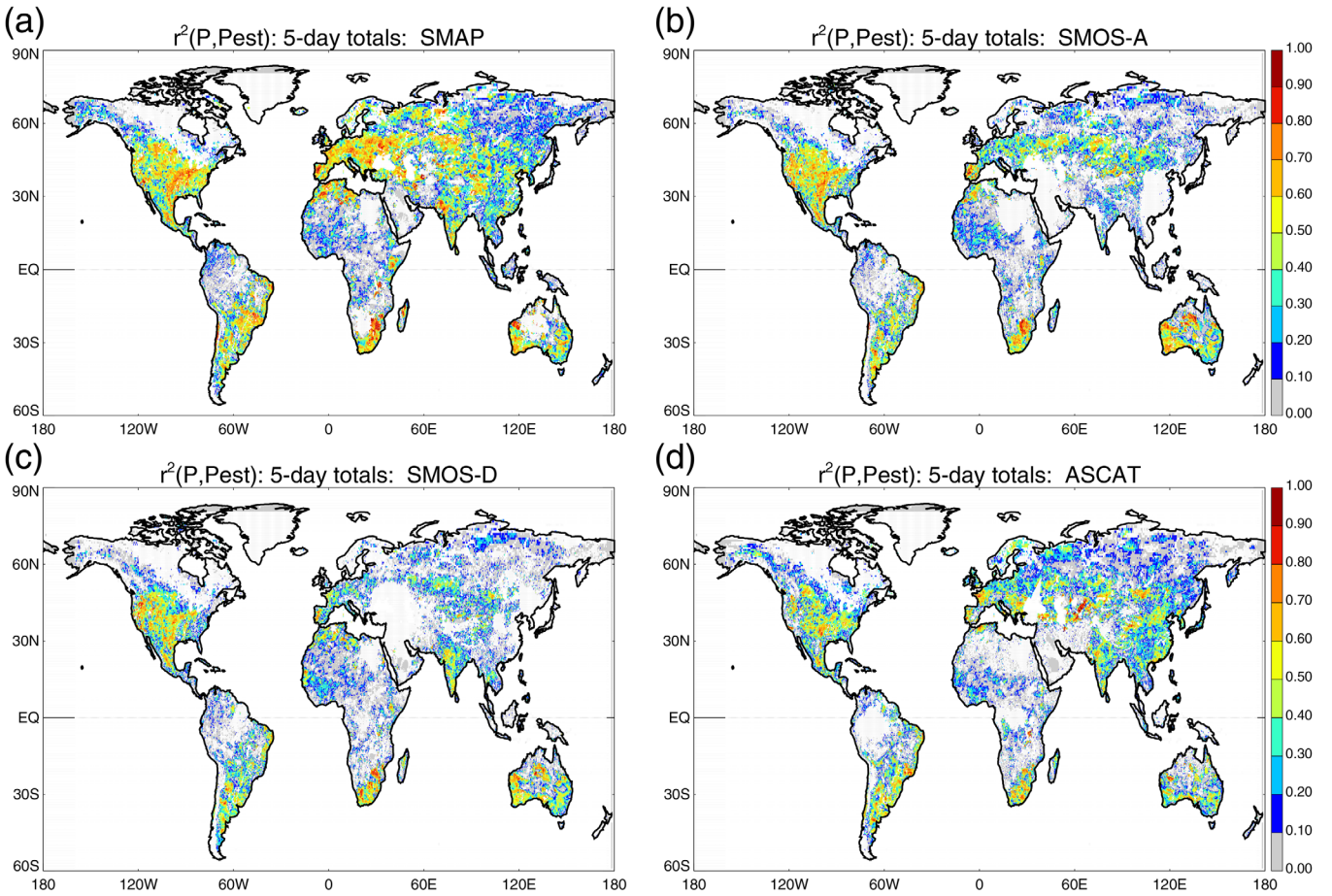


**Figure 4.** Square of the correlation coefficient ($r^2$) between 5 day precipitation totals from the CPCU rain gauge network and corresponding SM2RAIN-based precipitation estimates derived from (a) SMAP, (b) SMOS-A, (c) SMOS-D, and (d) ASCAT soil moisture retrievals. Gray coloring denotes correlations below 0.1; white coloring denotes locations for which correlations are undefined due to limitations in data availability. Considering $r^2$ calculations over the 23 5 day segments of the full validation period (covering mid-June through mid-October), $r^2$ values exceeding 0.13 are significantly different from zero at the 95% level, and those exceeding 0.23 are significantly different from zero at the 99% level; note, however, that these significance levels must be adjusted in a very small subset of locations (which varies with data set) for which retrievals cover only a fraction of the validation period.

continents. For SMAP, for example, the $r^2$ values in these regions often exceed 0.7—over much of the globe (∼24% of the globe with a gauge density of at least 1 gauge per 0.5° grid cell), the SMAP-based estimates "explain" 70% or more of the variance in the observed 5 day precipitation rates.

Of particular interest is the relative performance of the four data sets. Figures 4 and 5 indicate that of the four retrieval data sets examined, SMAP produces the most accurate precipitation estimates, followed by SMOS-A, with both performing better than SMOS-D and ASCAT. This relative performance is also apparent in the global averages of the skill levels shown in Figure 5: for ∼100 km averages, the average $r^2$ skill levels obtained over land areas for SMAP, SMOS-A, SMOS-D, and ASCAT are, respectively, 0.35, 0.29, 0.25, and 0.27. (Note that $r^2$ values are not computed in the whited out regions of the maps due to the presence of open water (exceeding a fraction of 0.05, according to SMAP estimates) or to other data limitations, such as those associated with RFI for SMOS. For a consistent comparison, the above global averages were computed across the same set of grid cells for each satellite data set—the set of cells holding a defined value for each data set.) Averaging instead over land areas with a gauge density of 1 gauge or more per 0.5° grid cell naturally gives higher (and more physically meaningful) averages: 0.58, 0.51, 0.43, and 0.42 for SMAP, SMOS-A, SMOS-D, and ASCAT, respectively.

We generalize further the relative skill levels of the different data sets and the impact of rain gauge density on this skill in Figure 6. For a given satellite retrieval data set, and for both the 36 km and aggregated ∼100 km resolutions, we compute the average of the precipitation estimation skill ($r^2$) over all land points having a gauge density within a stated range. Over 1000 values contribute to each average.

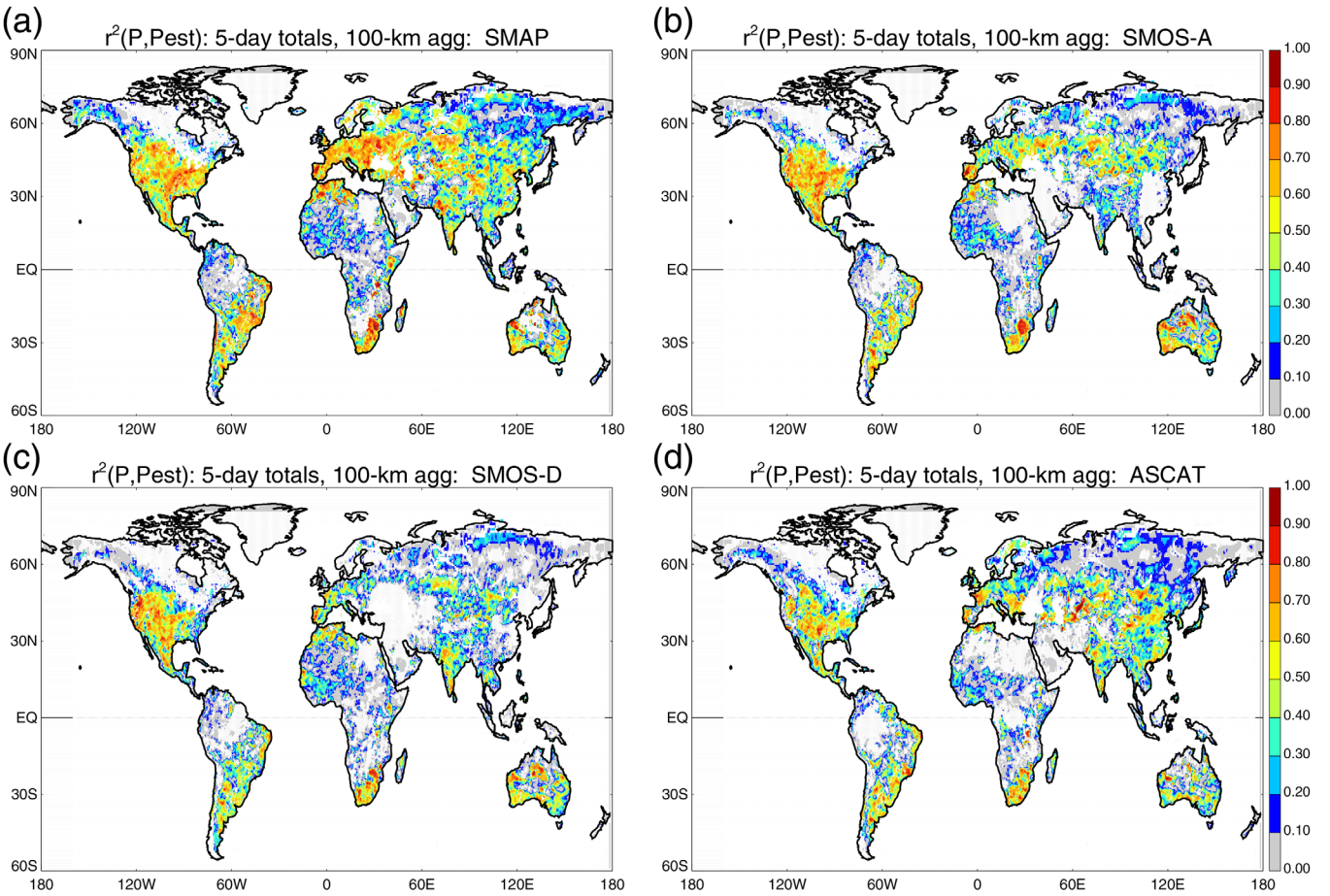


**Figure 5.** As in Figure 4, but for estimated and measured rain rates spatially aggregated to roughly a 1°×1° grid.

Two results are clearly evident from Figure 6. First, for all retrieval data sets, precipitation estimation skill increases with rain gauge density up to a density of about 1 gauge per 0.5° grid cell, after which it either grows less quickly with density (SMAP and SMOS-A) or plateaus to a roughly constant value (SMOS-D and ASCAT). Clearly, rain gauge density must be considered when evaluating the precipitation estimates. Second, the relative performance of the different retrieval data sets remains largely as noted above. SMAP provides the highest skill levels regardless of gauge density, followed by SMOS-A. SMOS-D and ASCAT perform similarly, with ASCAT performing slightly better at low rain gauge densities.

What causes these differences in precipitation estimation skill between the retrieval data sets? We can speculate that the differences are related to the inherent noise levels of the data sets. All soil moisture retrievals are subject to some noise, and by *differencing* two consecutive retrievals in (1), the impact of noise (particularly high frequency noise) on the accuracy of the precipitation estimates is amplified. In simple terms, greater amounts of noise must lead to reduced accuracy in precipitation estimation. Relative to the SMOS data, the SMAP data arguably have reduced noise and thus a greater potential for accurate precipitation estimation, given that the SMOS retrieval algorithm attempts to estimate multiple variables and given the emphasis on RFI mitigation techniques built into the SMAP system [*Entekhabi et al.*, 2010]. Given such considerations, the higher skill levels seen for SMAP make sense. It must be kept in mind, however, that for applications not as affected by high frequency noise, the SMAP and SMOS data sets have a presumably comparable usefulness.

Of the two SMOS data sets, SMOS-A is expected to be less noisy; *Lei et al.* [2015] demonstrate that, for most of the continental United States, SMOS-A retrievals are more accurate than SMOS-D retrievals. SMOS-A retrievals may have reduced noise due to the character of the vertical temperature profile in the soil at the time of the retrievals. The SMOS-A data were collected at 6 AM, whereas the SMOS-D data were collected at

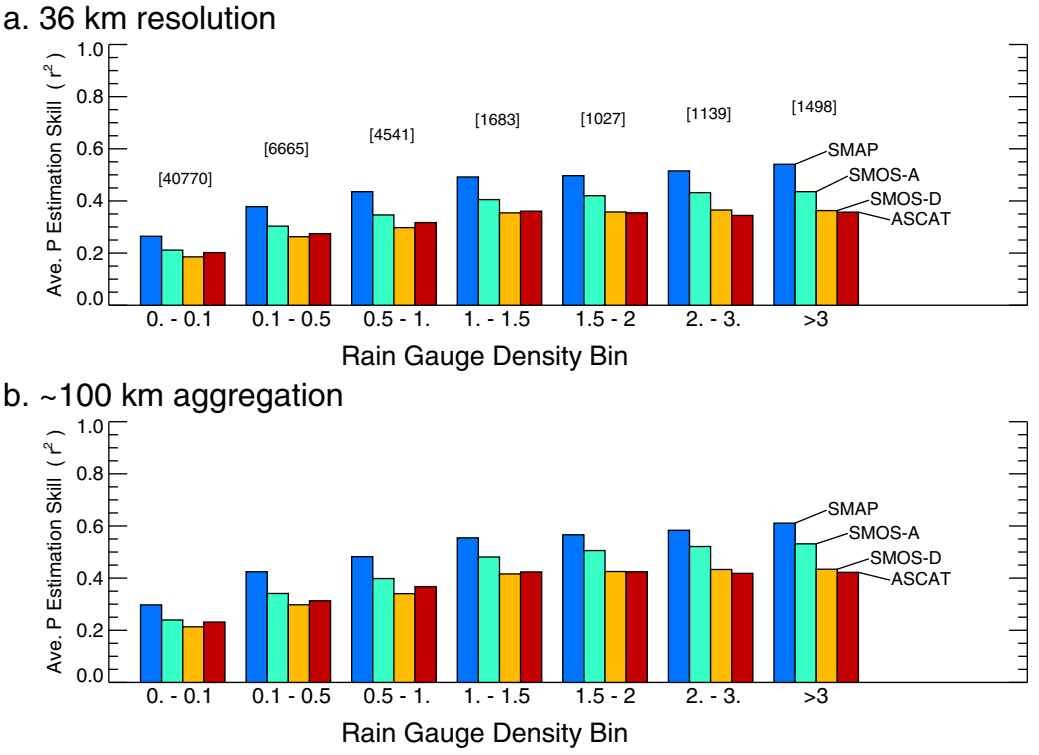


**Figure 6.** (a) Averages (across the globe, over grid cells holding data for each data set) of precipitation estimation skill ($r^2$) at the 36 km resolution for the four different data sets, binned according to rain gauge density (#gauges/0.5°×0.5° grid cell, as in Figure 2). That is, an $r^2$ value at a given location is included in an average if the local rain gauge density falls within the indicated range. (b) Same, but for the aggregated ~100 km resolution estimates shown in Figure 5. In the top plot, the number of grid cells contributing to a given binned average is provided in brackets above the histogram bars.

6 PM; various studies [e.g., *O'Neill et al*., 2014 (see their Figure 6)] suggest that at 6 AM, vertical temperature profiles in the soil, upon which retrieval algorithms are based, are roughly uniform, whereas at 6 PM, strong vertical gradients exist that can make soil moisture estimation more difficult. (Note, however, that at least one study [*Hornbuckle and England*, 2005] found the opposite: more vertical uniformity in the evening.) Regardless of reason, assuming (following *Lei et al*. [2015]) that SMOS-A retrievals are less noisy, the higher precipitation estimation accuracy found for SMOS-A relative to SMOS-D makes sense, though the SMOS-D estimates presumably also incur reduced $r^2$ values from increased inconsistency with the CPCU gauge measurement times.

Again, both SMAP and SMOS are L-band instruments and thereby see emissions from deeper into the soil than C-band instruments such as ASCAT (~5 cm versus ~2 cm). In the context of characterizing the connection between soil moisture and precipitation, the increased depth is an advantage, for at least two reasons. First, the greater depth can distinguish a greater range of precipitation inputs—while a 1 cm rainfall event and a 2 cm event may both saturate a dry 2 cm layer (given a 50% porosity), the two events will produce distinctly different levels of soil moisture increase for a 5 cm layer. Second, deeper layers are characterized by greater persistence [e.g., *Koster and Suarez*, 2001]; bare soil evaporation will reduce the average soil moisture content of a 2 cm layer more quickly than that of a 5 cm layer, and thus the latter can better retain information about a precipitation event if the event and the subsequent soil moisture retrieval are separated by, say, a couple of days. For these reasons, and because L-band measurements of emissions from the soil are less affected by the presence of vegetation than are C-band measurements, we expect the L-band instruments to perform better with the precipitation estimation algorithm. This expectation is borne out by the comparisons in Figures 4–6.

At this point, it is worth revisiting the findings of *Brocca et al*. [2014], who quantified precipitation estimation skill levels for the C-band instruments of ASCAT and AMSR-E and for a previous version of the SMOS data, using 5 day and 1° × 1° aggregates. Reprocessing the data examined by *Brocca et al*. [2014] over a

time period consistent with that used in this study (June to October, although for 2010–2011), we find precipitation skill levels for ASCAT (not shown) that are similar to those shown in Figure 5d, though with some regional differences. Interestingly, the ASCAT skill levels in the Sahel shown by *Brocca et al*. [2014] for all of 2010–2011 are better than those for any of the sensors in Figure 5, perhaps because the full 2010–2011 period includes the sharp soil moisture transition associated with the Sahelian monsoon, which falls outside of June to October. Reprocessing the *Brocca et al*. [2014] ASCAT results for June to October of 2010–2011 (not shown) significantly reduces Sahelian skill levels. Skill levels obtained for AMSR-E for June to October of 2010–2011 (not shown) are substantially lower than those for ASCAT and thus are substantially lower than those shown in Figure 5 for any of the sensors.

Curiously, the skill levels found by *Brocca et al*. [2014] for SMOS are substantially lower than those presented here. Presumably this reflects our use here of a more recent version of the SMOS data (we use SMUDP2 v620, whereas *Brocca et al*. [2014] used SMUDP2 v5.51) and more detailed quality control, using recently updated information—an indication that the reprocessing of such data sets, which is a standard part of such missions, can have a profoundly positive impact. Recall that the SMAP data used in this paper are from a beta release, suggesting the distinct possibility that future incarnations of the SMAP data could provide precipitation estimates of even higher accuracy.

## 4. Summary and Discussion

Application of the SM2RAIN algorithm to SMAP soil moisture retrievals produces time series of precipitation with significant levels of accuracy across much of the globe (Figures 4a and 5a). The average of the $r^2$ values for 5 day, $\sim 1^\circ \times 1^\circ$ accumulated precipitation estimates versus corresponding rain gauge observations is about 0.6 in parts of the globe for which the precipitation measurements used for validation are particularly reliable (Figure 6). These skill levels are indeed unprecedented for soil moisture-based precipitation estimation, being significantly higher than previously published values [*Brocca et al*., 2014]. Application of the algorithm to the latest SMOS data set (for ascending overpasses) produces slightly less accurate precipitation rates, and application to ASCAT data produces even lower accuracies. The relative levels of skill found for the retrieval data sets make sense in the context of their presumed relative levels of high frequency noise: SMAP data, due to built-in RFI corrections, are expected to be less noisy than SMOS data, and the L-band instruments (SMAP and SMOS) are expected to produce less noise than C-band instruments because they deal better with moderate levels of vegetation and because they see emissions from deeper in the soil, allowing a better discernment of different rainfall volumes.

One question, however, not fully addressed here is whether the use of ASCAT ascending data together with the descending data would have improved the skill levels produced for ASCAT. Because the ASCAT retrievals are based on a change detection algorithm, and because active products are less sensitive to land surface thermal conditions than passive products, soil temperature profiles are not a major issue for ASCAT, meaning that (in potential contrast to SMOS) ascending and descending ASCAT retrievals should have similar quality. When we reprocessed the 2010–2011 June to October ASCAT data examined by *Brocca et al*. [2014], which do include both ascending and descending data, we found skill levels (not shown) similar to those in Figure 5d, suggesting that use of the additional data would have had little effect. Still, the following caveat is worth mentioning: a definitive C-band analysis that includes 2015 ascending data have not yet been performed.

As illustrated in Figure 6, rain gauge density is an important consideration in the evaluation of the precipitation estimates. Our results indeed suggest that if precipitation rates were better measured in the ungauged areas, the skill levels obtained there would be higher. This has important implications. The high agreement in well-gauged areas suggests that retrieval-based precipitation estimates could be used for various applications there in lieu of gauge-based measurements, assuming enough observational precipitation data are available during a calibration period to scale the retrieval-based estimates to the proper magnitudes. If such scaling could be performed, then the retrieval-based precipitation estimates could themselves be used to drive, for example, a river routing or crop growth model. Now consider relatively ungauged regions (e.g., parts of the Sahel), for which the quality of the precipitation measurements is poor. Assuming that the retrievals have the same basic accuracy everywhere, and assuming that scaling factors obtained for well-gauged areas could be transferred to ungauged areas based on soil type and other considerations, our

results suggest that the retrieval-based precipitation estimates could be applied to great advantage in these areas—the estimates would arguably be better than gauge-based precipitation products.

This is, of course, an ambitious interpretation of the results. The retrieval-based precipitation estimates would presumably be poor in tropical forests (e.g., the Amazon) given known deficiencies of soil moisture retrievals in regions of dense vegetation. The retrievals may also be poorer in ungauged regions because model-based surface temperature estimates in these regions, a critical part of the retrieval algorithms (at least for SMOS and SMAP), may also be poor. Still, given that precipitation is generally more difficult to capture correctly than temperature, the interpretation is worth exploring with further study.

In any case, as noted in section 1, perhaps the greatest value of the soil moisture-based precipitation estimates lies in their potential combination with alternative precipitation estimates to produce a single, superior precipitation data set. This potential depends in large part on the degree to which the soil moisture-based estimates provide unique and complementary information about the temporal and spatial distributions of precipitation in nature. Devising an optimal strategy for combining the soil moisture-based estimates with those from, for example, gauge networks and satellite-based precipitation retrievals is beyond the scope of this paper; note, however, that relevant issues have been discussed in several recent studies [e.g., *Crow et al*., 2011; *Pellarin et al*., 2013; *Ciabatta et al*., 2015; *Zhan et al*., 2015]. Here we can address the complementarity of the information content by pointing to the strengths and weaknesses of each estimation approach.

Again, as noted in section 1, in situ gauge measurements, while providing direct (and thus high quality) measurements at gauge sites at high time resolution, are point measurements and do not necessarily capture well the precipitation that falls across large areas. Gauges are, in any case, sparse or wholly absent in many parts of the globe. Satellite-based precipitation measurements (e.g., from GPM) provide high temporal (e.g., half-hourly) and spatial resolution (e.g., 0.1°) data but to some degree are limited by both the "snapshot" nature of the different contributing measurements (thereby potentially missing rainfall amounts falling between the snapshots) and by difficulties, for example, in interpreting the relevant radiances over land.

The advantages and disadvantages of the soil moisture-based precipitation estimation approach are quite different. Disadvantages include a relatively coarser temporal resolution (2–3 days), as determined by the timespan between soil moisture retrievals. The estimates also necessarily miss any rainfall that: (i) runs directly off the surface, for example, during heavy storms or as encouraged by complex terrain (though as suggested by *Crow et al*. [2011], this impact may be minimal at the spatial scales considered here), or (ii) infiltrates quickly to deeper soil layers or evaporates quickly from the surface before the next soil moisture retrieval is captured. In addition, errors in soil moisture estimation at L-band and C-band are known to be large over dense vegetation and certain other surface types, meaning that the precipitation estimates in certain regions will be questionable. The advantages, however, of the soil moisture-based approach are potentially quite powerful. Relative to gauge measurements, the approach provides areally averaged estimates that span much more of the globe. Relative to direct satellite-based precipitation retrievals, the soil moisture-based estimates provide a time-integrated look at what happened between the soil moisture retrievals (akin to gauge measurements, but for large areas)—precipitation amounts falling between the "snapshots" of precipitation retrievals can be captured with the soil moisture-based estimation approach.

We emphasize again that it is presumably by combining approaches, emphasizing the strength of each one, that an optimal global precipitation data set can be constructed. This idea effectively underlies the aforementioned approaches of *Crow et al*. [2009, 2011], *Pellarin et al*. [2013], *Wanders et al*. [2015], and *Zhan et al*. [2015], and it is perhaps the best way to consider the SM2RAIN estimates examined here—not as a standalone precipitation data set but as a potential contributor to overall global precipitation estimation. The high skill levels shown in Figures 4–6, particularly for the L-band sensors, indicate that soil moisture retrievals do show significant promise for making such contributions.

**Acknowledgments**
Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). This work (mainly carried out at the NASA Goddard Space Flight Center) was supported by the NASA SMAP mission and the SMAP Science Team. Author Brocca appreciates support from the Italian Department of Civil Protection. Qing Liu and Clara Draper assisted with the processing of the data. SMAP data are available from https://nsidc.org/data/smap, SMOS data from https://smos-ds-02.eo.esa.int/oads/access/, and ASCAT data from http://www.eumetsat.int/website/home/index.htm. Precipitation data are available from ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB.

## References

Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, *4*, 1147–1167.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, *Geophys. Res. Lett.*, *40*, 853–858, doi:10.1002/grl.50173.

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *J. Geophys. Res. Atmos.*, *119*, 5128–5141, doi:10.1002/2014JD021489.

Brodzik, M. J., B. Billingsley, T. Haran, B. Raun, and M. H. Savoie (2012), EASE-Grid 2.0: Incremental but significant improvements for earth-gridded data sets, *ISPRS Int. J. Geo-Inf.*, *1*(1), 32–45.

Chan, S., and R. S. Dunbar (2015), SMAP level 2 passive soil moisture product specification document, *JPL Publ. JPL D-72547*, Jet Propul. Lab., Calif. Inst. of Technol., Pasadena, Calif.

Ciabatta, L., et al. (2015), Integration of satellite soil moisture and rainfall observations over the Italian territory. *J. Hydrometeorol.*, *16*, 1341–1355.

Crow, W. T., G. J. Huffman, R. Bindlish, and T. J. Jackson (2009), Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals, *J. Hydrometeorol.*, *10*, 199–212.

Crow, W. T., M. J. Van Den Berg, G. J. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resour. Res.*, *47*, W08521, doi:10.1029/2011WR010576.

De Lannoy, G. J. M, R. D. Koster, R. H. Reichle, S. P. P. Mahanama, and Q. Liu (2014), An updated treatment of soil texture and associated hydraulic properties in a global land modeling system, *J. Adv. Model. Earth Syst.*, *6*, 957–979, doi:10.1002/2014MS00330.

Entekhabi, D., et al. (2010), The Soil Moisture Active Passive (SMAP) mission, *Proc. IEEE*, *98*, 704–716.

Entekhabi, D., et al. (2014), SMAP handbook, *JPL Publ. JPL 400-1567*, 182 pp., Jet Propul. Lab., Pasadena, Calif.

Hornbuckle, B. K., and A. W. England (2005), Diurnal variation of vertical temperature gradients within a field of maize: Implications for satellite microwave radiometry, *IEEE Geosci. Remote Sens. Lett.*, *2*, 74–77.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The Global Precipitation Measurement Mission, *Bull. Am. Meteorol. Soc.*, *95*, 701–722.

Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, *78*, 5–20.

Huffman, G. J., et al. (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*, 38–55.

Jackson, T., et al. (2015), Calibration and validation for the L2/3_SM_P beta-release data products, version 2, *SMAP Project, JPL D-93981*, Jet Propul. Lab., Pasadena, Calif.

Kerr, Y. H., et al. (2010), The SMOS mission: New tool for monitoring key elements of the global water cycle, *Proc. IEEE*, *98*, 666–687, doi:10.1109/JPROC.2010.2043032.

Koster, R. D., and M. J. Suarez (2001), Soil moisture memory in climate models, *J. Hydrometeorol.*, *2*, 558–570.

Kummerow, C. D., D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. J. Munchak, and V. Petkovic (2015), The evolution of the Goddard Profiling Algorithm to a fully parametric scheme, *J. Atmos. Oceanic Technol.*, *32*, 2265–2280.

Lei, F., W. T. Crow, H. Shen, R. M. Parinussa, and T. H. Holmes (2015), The impact of local acquisition time on the accuracy of microwave surface soil moisture retrievals over the contiguous United States, *Remote Sens.*, *7*, 13,448–13,465, doi:10.3390/rs71013448.

Mahanama, S. P., R. D. Koster, G. K. Walker, L. L. Takacs, R. H. Reichle, G. De Lannoy, Q. Liu, B. Zhao, and M. J. Suarez (2015), Land boundary conditions for the Goddard Earth Observing System Model Version 5 (GEOS-5) Climate Modeling System—Recent updates and data file descriptions, *Technical Report Series on Global modeling and Data Assimilation*, vol. 39, NASA/TM-2015-104606, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Md.

O'Neill, P., S. Chan, E. Njoku, T. Jackson, and R. Bindlish (2014), Algorithm theoretical basis document level 2 & 3 soil moisture (passive) data products, *JPL D-66480*, Jet Propul. Lab., Calif. Inst. of Technol., Pasadena, Calif.

Pellarin, T., A. Ali, F. Chopin, I. Jobard, and J.-C. Bergès (2008), Using spaceborne surface soil moisture to constrain satellite precipitation estimates over West Africa, *Geophys. Res. Lett.*, *35*, L02813, doi:10.1029/2007GL032243.

Pellarin, T., S. Louvet, C. Gruhier, G. Quantin, and C. Legout (2013), A simple and effective method for correcting soil moisture and precipitation estimates using AMSR-E measurements, *Remote Sens. Environ.*, *136*, 28–36.

Wanders, N., M. Pan, and E. F. Wood (2015), Correction of real-time satellite precipitation with multi-sensor satellite observations of land surface variables, *Remote Sens. Environ.*, *160*, 206–221.

Wigneron J. P., et al. (2007), L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields, *Remote Sens. Environ.*, *107*, 639–655.

Xie, P., A. Yatagai, M. Chen, T. Hayasaka, Y. Fukushima, C. Liu, and S. Yang (2007), A gauge-based analysis of daily precipitation over East Asia, *J. Hydrometeorol.*, *8*, 607–626.

Zhan, W., M. Pan, N. Wanders, and E. F. Wood (2015), Correction of real-time satellite precipitation with satellite soil moisture observations, *Hydrol. Earth Syst. Sci.*, *19*, 4275–4291.